DOCKET NO: 248162US2SX

IN THE UNITED STATES PATENT & TRADEMARK OFFICE

IN RE APPLICATION OF	:	
YUTAKA KASHIHARA, ET AL.	:	EXAMINER: GUPTA, P.H.
SERIAL NO: 10/763,247	:	
FILED: JANUARY 26, 2004	:	GROUP ART UNIT: 2627
FOR: SIGNAL EVALUATION METHOD, INFORMATION RECORDING/REPRODUCING APPARATUS, INFORMATION REPRODUCING APPARATUS, AND INFORMATION RECORDING MEDIUM	:	

APPEAL BRIEF

COMMISSIONER FOR PATENTS
ALEXANDRIA, VIRGINIA 22313

SIR:

This is an appeal of the final Action (hereinafter "FA) mailed June 18, 2007, that presented a final rejection of Claims 1-16. A Notice of Appeal was timely filed with a one-month extension of time on October 18, 2007.

I. REAL PARTY IN INTEREST

The real parties in interest in this appeal are the Assignees KABUSHIKI KAISHA TOSHIBA and NEC Corporation.

II. RELATED APPEALS AND INTERFERENCES

Appellants, Appellants' legal representative, and the assignees are aware of no appeals which will directly affect or be directed affected by or have a bearing on the Board's decision in this appeal.

III. STATUS OF THE CLAIMS

Claims 1-16 are pending in this application. Claims 1-16 have been finally rejected and form the basis for this appeal. The attached claim appendix includes a clean copy of appealed Claims 1-16.

IV. STATUS OF THE AMENDMENTS

An Amendment after final was filed on September 17, 2007, to correct typographical errors in Claims 10-12 and 14-16. The Advisory Action mailed October 2, 2007, indicates that this Amendment of September 17, 2007, has been entered for purposes of appeal.

V. SUMMARY OF THE CLAIMED SUBJECT MATTER[1]

The subject matter of independent Claim 1 includes a signal evaluation method that evaluates a reproduction equalization signal reproduced from a recording medium by use of a PRML (partial response and maximum likelihood) discrimination method. This PRML discrimination method utilizes PR (partial response) characteristics, which correspond to the recording/reproducing characteristics as set forth in the Specification at page 5, lines 13-26, for example.

[1] It is Appellants' understanding that, under the rules of Practice before the Board of Patent Appeals and Interferences, 37 CFR §41.37(c) requires that a concise explanation of the subject matter recited in each independent claim be provided with reference to the specification by page and line numbers and to the drawings by reference characters. However, Appellants' compliance with such requirements anywhere in this document should in no way be interpreted as limiting the scope of the invention recited in all pending claims, but simply as non-limiting examples thereof.

Besides reciting that the signal evaluation method evaluates a reproduction equalization signal reproduced from a recording medium by use of a PRML, Claim 1 requires detecting matching between discrimination data and a plurality of predetermined bit pattern pairs of different groups. These pairs and groups are described in the specification at page 13, lines 5 -15, for example. Claim 1 also requires calculating a bit pattern and corresponding two ideal responses when the matching is detected, obtaining Euclidean distances between the two ideal responses and equalization reproduced signals, obtaining a difference between the Euclidean distances, obtaining a mean value and a standard deviation with respect to the difference between the Euclidean distances, and calculating a miss-discrimination probability F(0) of the predetermined bit pattern from the mean value and the standard deviation. Page 6 of the specification, at lines 1- 24, define the meaning of the terms “ideal responses” and “Euclidean distances,” while page 12, lines 8-27 explain these claimed steps relative to the exemplary steps A1-A6 of exemplary FIG. 4 considered with the above noted definitions and descriptions.

In addition, Claim 1 requires the final step of calculating a quality evaluation value of a reproduction signal based on the miss-discrimination probability F(0), an appearance probability of the predetermined bit pattern, and a Hamming distance between the predetermined bit pattern pairs as taught by step A7 of exemplary FIG. 4 and page 12, line 27 to page 13, line 1 that references the use of “formula (4).” This “formula (4)” is noted at page 10 line 10 of the specification and the terms used in this “formula (4)” are explained at page 10, lines 6-14.

Independent Claim 6 is directed to an apparatus used as one of an information recording/reproducing apparatus and an information reproducing apparatus that outputs reproduction signals reproduced from a recording medium by use of a PRML discrimination method, as described above relative to the specification at page 5, lines 13-26, for example.

Claim 6 recites an evaluation means (shown by 107 of Fig. 3, for example) including means for detecting matching between discrimination data and a plurality of predetermined bit pattern pairs of different groups (shown by 203 and 202) of Fig. 3, for example), means for calculating a bit pattern and corresponding two ideal responses when the matching is detected (shown by 204 of Fig. 3, for example).

As explained in the specification at page 12, lines 8- 27, for example, the evaluation calculating unit 204 of Fig.3 also functions as the Claim 6 means for obtaining Euclidean distances, means for obtaining a difference between the Euclidean distances, means for obtaining a mean value and a standard deviation with respect to the difference between the Euclidean distances, means for calculating a miss-discrimination probability F(0) of the predetermined bit pattern from the mean value and the standard deviation, and means for calculating a quality evaluation value of a reproduction signal based on the miss-discrimination probability F(0), an appearance probability of the predetermined bit pattern, and a Hamming distance between the predetermined bit pattern pair as it calculates the values of D expressed by formula (1) (in the specification at page 9, line 15) using equalization signals S, the ideal signal PT of pattern T, and ideal signal PF of pattern F (also described on page 9 of the specification at lines 6-14); the mean value and a standard deviation with respect to a plurality of D values; the value of F(0) expressed by formula (3) (in the specification at page 10) after an ample number of pieces of data are processed and the data acquisition ends, and the quality evaluation value of a reproduction signal based on the miss-discrimination probability F(0), an appearance probability of the predetermined bit pattern, and a Hamming distance between the predetermined bit pattern pairs as expressed by formula (4) (also in the specification at page 10).

The subject matter of independent Claim 10 as to an information recording medium from which reproduction signals are reproduced by use of a PRML discrimination method.

This PRML discrimination method is noted above as to Claims 1 and 6 to utilize PR (partial response) characteristics, which correspond to the recording/reproducing characteristics set forth in the Specification at page 5, lines 13-26, for example.

Claim 10 also requires that the reproduction signals be evaluated based on an evaluation value obtained by detecting matching between discrimination data and a plurality of predetermined bit pattern pairs of different groups. These pairs and groups are described in the specification at page 13, lines 5 -15, for example. Claim 10 further requires calculating a bit pattern and corresponding two ideal responses when the matching is detected, obtaining Euclidean distances between the two ideal responses and equalization reproduced signals, obtaining a difference between the Euclidean distances, obtaining a mean value and a standard deviation with respect to the difference between the Euclidean distances, and calculating a miss-discrimination probability F(0) of the predetermined bit pattern from the mean value and the standard deviation. Page 6 of the specification, at lines 1- 24, define the meaning of the terms "ideal responses" and "Euclidean distances," while page 12, lines 8-27 explain these claimed steps relative to the exemplary steps A1-A6 of exemplary FIG. 4 considered with the above noted definitions and descriptions.

In addition, Claim 10 requires the step of calculating a quality evaluation value of a reproduction signal based on the miss-discrimination probability F(0), an appearance probability of the predetermined bit pattern, and a Hamming distance between the predetermined bit pattern pairs as taught by step A7 of exemplary FIG. 4 and page 12, line 27 to page 13, line 1 that references the use of "formula (4)." This "formula (4)" is noted at page 10 line 10 of the specification and the terms used in this "formula (4)" are explained at page 10, lines 6-14.

Lastly, Claim 10 requires that the information recording medium satisfies a requirement that the evaluation value is not more than 1×10^{-3} as noted at exemplary page 16, lines 16-23, of the specification.

Independent Claim 14 is similar to independent Claim 10 in that it also to an information recording medium from which reproduction signals are reproduced by use of a PRML discrimination method. This PRML discrimination method is noted above as to Claims 1, 6, and 10 to utilize PR (partial response) characteristics, which correspond to the recording/reproducing characteristics set forth in the Specification at page 5, lines 13-26, for example.

Claim 14 also requires that the reproduction signals be evaluated based on an evaluation value obtained by detecting matching between discrimination data and a plurality of predetermined bit pattern pairs of different groups. These pairs and groups are described in the specification at page 13, lines 5 -15, for example. Claim 14 further requires calculating a bit pattern and corresponding two ideal responses when the matching is detected, obtaining Euclidean distances between the two ideal responses and equalization reproduced signals, obtaining a difference between the Euclidean distances, obtaining a mean value and a standard deviation with respect to the difference between the Euclidean distances, and calculating a miss-discrimination probability F(0) of the predetermined bit pattern from the mean value and the standard deviation. Page 6 of the specification, at lines 1- 24, define the meaning of the terms "ideal responses" and "Euclidean distances," while page 12, lines 8-27 explain these claimed steps relative to the exemplary steps A1-A6 of exemplary FIG. 4 considered with the above noted definitions and descriptions.

In addition, Claim 14 requires the step of calculating a quality evaluation value of a reproduction signal based on the miss-discrimination probability F(0), an appearance probability of the predetermined bit pattern, and a Hamming distance between the

predetermined bit pattern pairs as taught by step A7 of exemplary FIG. 4 and page 12, line 27 to page 13, line 1 that references the use of "formula (4)." This "formula (4)" is noted at page 10 line 10 of the specification and the terms used in this "formula (4)" are explained at page 10, lines 6-14.

Lastly, Claim 14 requires that the information recording medium satisfies a requirement that the evaluation value is not more than 1×10^{-5} as noted at exemplary page 16, line 24 to page 17, line 3, of the specification.

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

Claims 1-4, 6-8, 10 and 14 have been finally rejected as being unpatentable over Okumura et al. (U.S. Patent Publication 2005/0193318, hereinafter Okumura) in view of Hagenauer et al. (U.S. Patent 5,181,209 hereinafter Hagenauer); and Claims 5, 9, 11-13 and 15-16 have been finally rejected as being unpatentable over Okumura.

VII. ARGUMENT

A. The rejection of Claims 1-4, 6-8, 10 and 14 as being unpatentable over Okumura in view of Hagenauer

1. Independent Claim 1

Paragraph 4 on page 8 of the "FA" asserts that paragraph [0120] of Okumura describes a path metric difference which is used as a "Hamming distance." The "FA" also asserts that Applicants' claimed appearance probability is equivalent to the probabilities described in paragraph [0121] of Okumura. Applicants note these characterizations of Okumura to be clear errors.

Okumura describes a device configured to adaptively equalize a waveform of a Viterbi-decodable input signal pattern. The device includes an FIR filter for generating an equalized signal pattern; a Viterbi decoder for detecting a path metric difference between a correct path and error path; a target value register for setting a target value for the path metric difference, and a tap coefficients update circuit for adapting the equalization according to an error of the detected path metric difference from the target value, see paragraph [0043] of Okumura.

Paragraph [0120] of Okumura describes Figure 5A as a histogram of a path metric difference from a totally noise free ideal waveform. The path metric difference takes discrete values (ideal values). The ideal values vary because the path metric difference of an error path emanating from the same state and terminating on the same state differs from one bit pattern to another. Figure 6 of Okumura shows a relationship between ideal values of the path metric difference and bit patterns corresponding to the ideal values. Paragraph [0173] of Okumura describes that, since the waveform interference width is 4T, a trellis diagram will show six states. A minimum ideal value for a path metric difference is given by 8 bit patterns by which the error path merges with the correct path with the least number of state transitions.

The discussion of the background in Okumura teaches that jitter has often been used to evaluate reproduced signal quality on an optical disc. Okumura further notes that PRML is a data detection method for realizing higher density storage. However, with PRML, jitter is not suitable as an evaluation value. Instead, a bit error rate that has been obtained by PRML is used as the evaluation value. Also, this PRML evaluation value requires a large number of sample bits for a measurement, and defects on the disk tend to influence the evaluation accuracy. Paragraphes [0018-0019] of Okumura note an evaluation method, called SAM (Sequenced Amplitude Margin), has been proposed to address these issues.

SAM is described FIGS. 21 and 22 of Okumura, where a reproduced signal of a bit pattern that has been recorded on the basis of d=1 (1, 7) RLL (Run Length Limited) Coding is decoded in PRML, in accordance with PR (1, 2, 1) properties. As shown in FIG. 21, a reproduced waveform in accordance with PR(1,2,1) properties has an ideal 1T mark free from any distortion or noise and has a 1:2:1 level ratio of samples for a channel clock. For a reproduced waveform from a 2T or more mark, a level ratio is obtainable from the superimposition of the reproduced waveform from a 1T mark. . For example, the sample level ratio is 1:3:3:1 for the 2T mark, 1:3:4:3:1 for the 3T mark, and 1:3:4:4:3:1 for the 4T mark. An ideal reproduced waveform can be assumed for any given bit pattern. Paragraphes [0020-0021] of Okumura teach five ideal sample levels (ideal sample levels): 0, 1, 2, 3, and 4.

In Okumura, Viterbi decoding is adopted for PRML decoding. The Viterbi decoding is described as follows with reference to a trellis diagram shown in FIG. 22. In FIG. 22, S(00), S(01), S(10), and S(11) each represents a different state: for example, the state S(00) means a 0 previous bit and a 0 current bit. A line linking a state to the other is termed a "branch," which represents a state transition: for example, a branch of S(00)-S(01) represents a "001" bit pattern. Here, the S(01)-S(10) and S(10)-S(01) branches are missing from the diagram. This is because the 010 and 101 bit patterns cannot occur due to the d=1 (1,7) RLL. Each letter of α to ξ is allocated to each branch as an identifier, and an ideal waveform level expected at each state transition follows the identifier. In PR (1, 2, 1), an ideal waveform level is determined by three successive bits: v0. v1, and v2, and a value of the ideal waveform level is calculated by v0+2v1+v2. For example, when a represents a "000" bit pattern, the ideal level is 0, and when β represents a "100" bit pattern, the ideal level is 1, see paragraphes [0022-0023] of Okumura.

In the trellis diagram, a "path" is formed by connecting continuous branches between the states. To consider all the paths generated after transiting from any one state to another means to consider all the possible bit patterns. The most likely path, or the "correct path," can be determined by comparing the waveform actually reproduced from the optical disc with every ideal waveform derived from the paths to find the ideal waveform that is the "closest" to the reproduced waveform (that is, the one with the least Euclidean distance from the reproduced waveform.), see Okumura at paragraph [0024].

However, contrary to the FA at page 3, lines 7-12, Okumura does not teach or suggest Applicants' claimed step of calculating a quality evaluation value of a reproduction signal based on the miss-discrimination probability F(0), an appearance probability of the predetermined bit pattern, and a Hamming distance between the predetermined bit pattern pairs at least because a path metric distance is not a Hamming distance.

By way of background, the following analysis is provided to explain the differences between Applicants' claimed Hamming Distance and the Path Metric Difference of Okumura. For example, assume that the predetermined bit pattern pair is as follows: Pattern A: 11000; and Pattern B: 11100. The "Hamming Distance" refers to the number of different bits between two patterns. This meaning of "Hamming Distance" is well recognized in the art, see U.S. Patent No. 6 226,640 at col. 1, lines 39-44, included in the Evidence Appendix, for example. In this case, the Hamming Distance between Patterns A and B is 1. The path metric of Okumura provides a different result as seen by the following.

Assume that PR (1, 2, 1) is used in PR class (in accordance with the reference). In this case, an ideal response to each of Patterns A and B is as follows: Ideal Response A: 13310; and Ideal Response B: 13431. Assume that a real equalization reproduced signal has the following contents: Real Equalization Reproduced Signal: [0.8 2.5 3.5 2.5 0.5].

In this case, the Euclid distance between the real equalization reproduced signal and the ideal response A, and the Euclid distance between the real equalization reproduced signal and the ideal response B, are expressed by the following formulas:

Euclid Distance A= $\sqrt{\{(1-0.8)^2+(3-2.5)^2+(3-2.5)^2+(1-2.5)^2+(0-0.5)^2\}}=\sqrt{3.04}$

Euclid Distance B= $\sqrt{\{(1-0.8)^2+(3-2.5)^2+(4-3.5)^2+(3-2.5)^2+(1-0.5)^2\}}=\sqrt{1.04}$

In this case, the square of the Path Metric Difference is calculated as 3.04-1.04=2.00. Thus, contrary to the Official Action the Hamming Distance and the Path Metric Difference are clearly different. Thus, contrary to the Official Action, Okumura does not teach the use Applicants' claimed Hamming distance as to the Claim 1 required calculating a quality evaluation value of a reproduction signal.

Furthermore, the evaluation value calculated in the reference can express the probability that Pattern B is misidentified as Pattern A, but is insufficient to express to what extent the bit error rate is affected. In order to express to what extent the bit error rate is affected by the misidentification of Pattern B as Pattern A, the occurrence probability of Pattern B and the number of bits that become bit errors as a result of one misidentification (Hamming Distance) are further required. The present invention enables the evaluation of the degree of effect on the bit error rate by using occurrence probability of a predetermined pattern and the Hamming Distance. No such benefit is possible with Okumura.

Furthermore, and as acknowledged by the Official Action, Okumura does not disclose or suggest calculating a miss-discrimination probability or calculating a quality evaluation value based on the miss-discrimination probability. To cure this deficiency, the Official Action applies Hagenauer.

Hagenauer describes a method and device for generalizing a Viterbi algorithm in such a way that the Viterbi algorithm produces analog, i.e., soft decisions. Figure 1 of Hagenauer

shows a Viterbi detector that provides estimates for a symbol sequence by processing the received symbol sequence in an MAP or Viterbi detector. Because the estimated value is not always correct, a conditional probability density function is provided which describes the estimated error. However, the estimated error of Hagenauer is not equivalent to Applicants' claimed miss-discrimination probability. Applicants' claimed miss-discrimination probability F(0) is defined in equation 3 on page 10 of Applicants' specification. This probability is the probability of a mistaken recognition of true as false. The Viterbi decoder does not calculate any type of miss-discrimination probability (i.e., the probability of a mistaken recognition of true as false), let alone the probability shown in equation 3.

Furthermore, like Okumura, Hagenauer does not disclose or suggest calculating a quality evaluation value of a reproduction signal based on a Hamming distance between the predetermined bit pattern pairs.

In summary, assuming *arguendo* that Okumura teaches calculating a quality evaluation value of a reproduction signal based on

- an appearance probability of the predetermined bit pattern, and
- a path metric difference,

Okumara does not teach or suggest calculating a quality evaluation value of a reproduction signal based on

- the miss-discrimination probability F(0), and
- a Hamming distance between the predetermined bit pattern pairs.

Hagenauer does not cure the deficiencies of Okumara.

Consequently, the rejection of method Claim 1 should be reversed.

2. Independent Claim 6

Independent Claim 6 is directed to an apparatus used as one of an information recording/reproducing apparatus and an information reproducing apparatus that outputs reproduction signals reproduced from a recording medium and requires the means plus function recitations that correspond to the steps recited by independent Claims 1.

Accordingly, even again assuming *arguendo* that Okumura teaches a means for calculating a quality evaluation value of a reproduction signal based on

- an appearance probability of the predetermined bit pattern, and
- a path metric difference,

Okumara does not teach or suggest means for calculating a quality evaluation value of a reproduction signal based on

- the miss-discrimination probability F(0), and
- a Hamming distance between the predetermined bit pattern pairs.

Again, Hagenauer does not cure the deficiencies of Okumara.

Consequently, the rejection of apparatus Claim 6 should also be reversed.

3. Independent Claim 10

The subject matter of independent Claim 10 as to an information recording medium from which reproduction signals are reproduced by use of the method of Claim 1 clearly defines over the combination of Okumara and Hagenauer for at least the same reasons as Claim 1 does. In addition, Claim 10 adds the requirement that the information recording medium must satisfy a requirement that the evaluation value is not more than 1×10^{-3}. The "FA" (at page 3, lines 13-14) points to the bit error rates of paragraphs [0147] and [0148] of Okumara. However, Claim 10 requires that "the evaluation value is not more than 1 × 10-

3"and this is a different parameter as compared to the relied upon Okumara bit error rates of paragraphs [0147] and [0148]. Again, Hagenauer does not cure the deficiencies of Okumara.

Consequently, the rejection of information recording medium Claim 10 should also be reversed.

4. Independent Claim 14

The subject matter of independent Claim 14 as to an information recording medium from which reproduction signals are reproduced by use of the method of Claim 1 clearly defines over the combination of Okumara and Hagenauer for at least the same reasons as Claim 1 does. In addition, Claim 14 adds the requirement that the information recording medium must satisfy a requirement that the evaluation value is not more than 1×10^{-5}. The "FA" (at page 3, lines 13-14) again points to the bit error rates of paragraphs [0147] and [0148] of Okumara. However, Claim 14 requires that "the evaluation value is not more than 1×10^{-5}"and this is a different parameter as compared to the relied upon Okumara bit error rates of paragraphs [0147] and [0148]. Again, Hagenauer does not cure the deficiencies of Okumara.

Consequently, the rejection of information recording medium Claim 14 should also be reversed.

B. The rejection of Claims 5, 9, 11-13 and 15-16 as being unpatentable over Okumura

Claims 5, 9, 11-13 and 15-16 are all dependent claims, with Claim 5 depending on Claims 1-4 and 9 depending from Claims 6 and 7, Claims 11-13 depending from Claim 10, and Claims 15-16 depending from Claim 14. The "FA" admits, however, that Okumura does not teach the step (and means) for calculating "a miss-discrimination probability F(0) of the

predetermined bit pattern from the mean value and the standard deviation" or for "calculating a quality evaluation value of a reproduction signal based on the miss-discrimination probability F(0)" as to the subject matter of parent Claims 1-4, 6, 7, 10, and 14 at page 3, lines 14-18. Whether or not Hagenauer can be said to cure these deficiencies of Okumura is not material because the stated rejection only relies upon Okumura, and Okumura does not contain the teachings of Hagenauer.

Consequently, the rejection of Claims 5, 9, 11-13 and 15-16 over Okumura alone should also be reversed.

CONCLUSION

For all the reasons noted above, the rejections applied to Claims 1-16 should be reversed as being clearly improper

Respectfully Submitted,

OBLON, SPIVAK, McCLELLAND,
MAIER & NEUSTADT, P.C.

Eckhard H. Kuesters
Registration No. 28,870
Attorney of Record

Raymond F. Cardillo, Jr.
Registration No. 40,440

Customer Number
22850
Tel. (703) 413-3000
Fax. (703) 413-2220
(OSMMN 10/01)

EHK:RFC:jmp

VII. CLAIMS APPENDIX

1. A signal evaluation method configured to evaluate a reproduction equalization signal reproduced from a recording medium by use of a PRML (partial response and maximum likelihood) discrimination method, said method comprising the steps of:

detecting matching between discrimination data and a plurality of predetermined bit pattern pairs of different groups;

calculating a bit pattern and corresponding two ideal responses when the matching is detected;

obtaining Euclidean distances between the two ideal responses and equalization reproduced signals;

obtaining a difference between the Euclidean distances;

obtaining a mean value and a standard deviation with respect to the difference between the Euclidean distances;

calculating a miss-discrimination probability F(0) of the predetermined bit pattern from the mean value and the standard deviation; and

calculating a quality evaluation value of a reproduction signal based on the miss-discrimination probability F(0), an appearance probability of the predetermined bit pattern, and a Hamming distance between the predetermined bit pattern pairs.

2. A signal evaluation method according to claim 1, wherein said quality evaluation signal is used as a first evaluation value, a target signal is calculated based on a predetermined data sequence and a predetermined partial response characteristic, an equalization error representing a difference in reproduction equalization signals is calculated in each clock period, a second evaluation value based on the autocorrelation of said equalization error is used as an evaluation value for evaluating the signal quality, and

said first evaluation value and said second evaluation value are used in combination to obtain final evaluation.

3. A signal evaluation method according to claim 2, wherein the final evaluation is made based on the first evaluation value, the second evaluation value, and a third evaluation value, the third evaluation value being provided by an error correction decoder and attributable mainly to a medium defect.

4. A signal evaluation method according to claim 1, wherein said quality evaluation value is used as a first evaluation value, and the final evaluation is made based on the first evaluation value and a third evaluation value, the third evaluation value being provided by an error correction decoder and attributable mainly to a medium defect.

5. A signal evaluation method according to any one of claims 1, 2, 3 and 4, wherein the evaluation value is calculated by use of equalization signals corresponding to 100,000 channel bits or more.

6. An apparatus used as one of an information recording/reproducing apparatus and an information reproducing apparatus and outputting reproduction signals reproduced from a recording medium by use of a PRML (partial response and maximum likelihood) discrimination method, said apparatus comprising signal reproduction evaluation means including:

means for detecting matching between discrimination data and a plurality of predetermined bit pattern pairs of different groups;

means for calculating a bit pattern and corresponding two ideal responses when the matching is detected;

means for obtaining Euclidean distances between the two ideal responses and equalization reproduced signals;

means for obtaining a difference between the Euclidean distances;

means for obtaining a mean value and a standard deviation with respect to the difference between the Euclidean distances;

means for calculating a miss-discrimination probability F(0) of the predetermined bit pattern from the mean value and the standard deviation; and

means for calculating a quality evaluation value of a reproduction signal based on the miss-discrimination probability F(0), an appearance probability of the predetermined bit pattern, and a Hamming distance between the predetermined bit pattern pairs.

7. An apparatus according to claim 6, further comprising:

means for adjusting a recording waveform by use of a value calculated based on the mean value and the standard deviation.

8. An apparatus used as one of an information recording/reproducing apparatus and an information reproducing apparatus and configured to produce an evaluation value by use of a signal evaluation method described in any one of claims 1, 2, 3, and 4, said apparatus comprising means for performing at least one of: adjustment of a recording waveform; an offset adjustment of a reproduction signal; gain adjustment; adjustment of an equalization coefficient; tracking control; focusing control; tilting control; and the adjustment of a spherical aberration.

9. An apparatus according to any one of claims 6 and 7, wherein the evaluation value is calculated by use of equalization signals corresponding to 100,000 channel bits or more.

10. An information recording medium from which reproduction signals are reproduced by use of a PRML (partial response and maximum likelihood) discrimination method, the reproduction signals being evaluated based on an evaluation value obtained by:

detecting matching between discrimination data and a plurality of predetermined bit pattern pairs of different groups;

calculating a bit pattern and corresponding two ideal responses when the matching is detected;

obtaining Euclidean distances between the two ideal responses and equalization reproduced signals;

obtaining a difference between the Euclidean distances;

obtaining a mean value and a standard deviation with respect to the difference between the Euclidean distances;

calculating a miss-discrimination probability F(0) of the predetermined bit pattern from the mean value and the standard deviation; and

calculating a quality evaluation value of a reproduction signal based on the miss-discrimination probability F(0), an appearance probability of the predetermined bit pattern, and a Hamming distance between the predetermined bit pattern pairs,

said information recording medium satisfying a requirement that the evaluation value is not more than 1×10^{-3}.

11. An information recording medium according to claim 10, wherein said quality evaluation signal is used as a first evaluation value, a target signal is calculated based on a

predetermined data sequence and a predetermined partial response characteristic, an equalization error representing a difference in reproduction equalization signals is calculated in each clock period, a second evaluation value based on the autocorrelation of the equalization error is used as an evaluation value for evaluating the signal quality, and said first evaluation value and said second evaluation value are used in combination to obtain final evaluation,

said information recording medium satisfying a requirement that the first evaluation value is not more than 1×10^{-3} and the second evaluation value is not less than 12.

12. A recording information medium according to claim 11, wherein the final evaluation is made based on the first evaluation value, the second evaluation value and a third evaluation value, the third evaluation value being provided by an error correction decoder, which performs error correction with respect to the reproduction signals, and attributable mainly to a medium defect,

said information recording medium satisfying a requirement that the first evaluation value is not more than 1×10^{-3}, the second evaluation value is not less than 12, and the third evaluation value is not more than 280 for 8 ECC consecutive blocks.

13. An information recording medium according to claim 10, wherein said quality evaluation signal is used as a first evaluation value, a target signal is calculated based on a predetermined data sequence and a predetermined partial response characteristic, an equalization error representing a difference in reproduction equalization signals is calculated in each clock period, a second evaluation value based on the autocorrelation of the equalization error is used as an evaluation value for evaluating the signal quality, and said

first evaluation value and said second evaluation value are used in combination to obtain final evaluation,

said information recording medium satisfying a requirement that the second evaluation value is not less than 15.

14. An information recording medium from which reproduction signals are reproduced by use of a PRML (partial response and maximum likelihood) discrimination method, the reproduction signals being evaluated based on an evaluation value obtained by:

detecting matching between discrimination data and a plurality of predetermined bit pattern pairs of different groups;

calculating a bit pattern and corresponding two ideal responses when the matching is detected;

obtaining Euclidean distances between the two ideal responses and equalization reproduced signals;

obtaining a difference between the Euclidean distances;

obtaining a mean value and a standard deviation with respect to the difference between the Euclidean distances;

calculating a miss-discrimination probability F(0) of the predetermined bit pattern from the mean value and the standard deviation; and

calculating a quality evaluation value of a reproduction signal based on the miss-discrimination probability F(0), an appearance probability of the predetermined bit pattern, and a Hamming distance between the predetermined bit pattern pairs,

said information recording medium satisfying a requirement that the evaluation value is not more than 1×10^{-5}.

15. An information recording medium according to claim 14, wherein said quality evaluation signal is used as a first evaluation value, a target signal is calculated based on a predetermined data sequence and a predetermined partial response characteristic, an equalization error representing a difference in reproduction equalization signals is calculated in each clock period, a second evaluation value based on the autocorrelation of the equalization error is used as an evaluation value for evaluating the signal quality, and said first evaluation value and said second evaluation value are used in combination to obtain final evaluation,

said information recording medium satisfying a requirement that the first evaluation value is not more than 1×10^{-5} and the second evaluation value is not less than 15.

16. A recording information medium according to claim 15, wherein the final evaluation is made based on the first evaluation value, the second evaluation value and a third evaluation value, the third evaluation value being provided by an error correction decoder, which performs error correction with respect to the reproduction signals, and attributable mainly to a medium defect,

said information recording medium satisfying a requirement that the first evaluation value is not more than 1×10^{-5}, the second evaluation value is not less than 15, and the third evaluation value is not more than 280 for 8 consecutive ECC blocks.

IX. EVIDENCE APPENDIX

U.S. Patent No. 6,226,640, Ostrovsky et al., May 1, 2001, attached.

US006226640B1

(12) United States Patent
Ostrovsky et al.

(10) **Patent No.: US 6,226,640 B1**
(45) **Date of Patent: May 1, 2001**

(54) **METHOD FOR DETERMINING APPROXIMATE HAMMING DISTANCE AND APPROXIMATE NEAREST NEIGHBORS OF A QUERY**

(75) Inventors: **Rafail Ostrovsky**, Secaucus, NJ (US); **Yuval Rabani**, Haifa (IL)

(73) Assignee: **Telecordia Technologies, Inc.**, Morristown, NJ (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/193,207**

(22) Filed: **Nov. 17, 1998**

Related U.S. Application Data

(60) Provisional application No. 60/066,936, filed on Nov. 17, 1997.

(51) **Int. Cl.**[7] **G06F 17/30**
(52) **U.S. Cl.** **707/5**
(58) **Field of Search** 707/5, 1–4

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,053,871	*	10/1977	Vidalin et al.	340/4
4,084,260	*	4/1978	Fleming et al.	707/3
5,855,018	*	12/1998	Chor et al.	707/9
5,870,754	*	2/1999	Dimitrova et al.	707/104
5,890,151	*	3/1999	Agrawal et al.	707/5

OTHER PUBLICATIONS

Balakirsky, V. B. "Hashing of Databases Based on Indirect Observation of Hamming Distances," Information Theory, IEEE Transactions on, Mar. 1996, vol. 42, Iss. 2, pp. 664–671.*

Ouyang, Y. C. et al. "Neural Network Based Retrieval Issue on Protoype Database Systems,"Oct. 1991 vol. 3, lines 1493–1497.*

J. Kleinberg, "Two Algorithms for Nearest–Neighbor Search in High Dimensions", Proceedings of the 29th Symposium of Theory of Computing, pp. 599–608, 1997.

P. Indyk and R. Motwani, "Approximate Nearest Neighbors: Towards Removing the Curse of Dimensionality", Proceedings of 30th Symposium of Theory of Computing, pp. 604–613, 1998.

* cited by examiner

Primary Examiner—Jack Choules
(74) *Attorney, Agent, or Firm*—Joseph Giordano

(57) **ABSTRACT**

A method and system identify in a database one or more data entries that are the nearest neighbors of a query. The database prebuilds a first set of strings by probabilistically selecting values of respective bits in each of the first set of strings based on a probability that depends on a first hamming distance. Based on the first set of strings, the database predetermines the trace values of each data entry in the database, respectively, and stores the predetermined trace values as entries in a trace table. For each trace value entry, the database identifies the data entries whose trace values are within a second hamming distance of the trace value entry, and stores the addresses of the identified data entries in the trace value entry. When the database receives a query, by identifying the trace value entry in the trace table that match the trace value of the query, the database identifies the data entries that are within the first hamming distance of the query. In addition, a method and system estimate the hamming distance between two strings in a network.

17 Claims, 7 Drawing Sheets

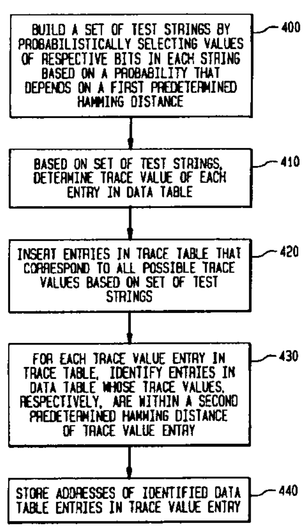


FIG. 1

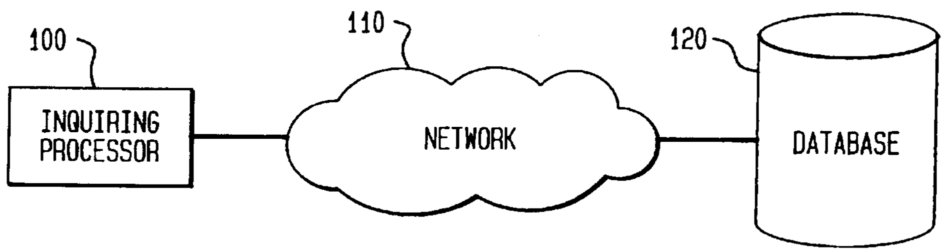


FIG. 2

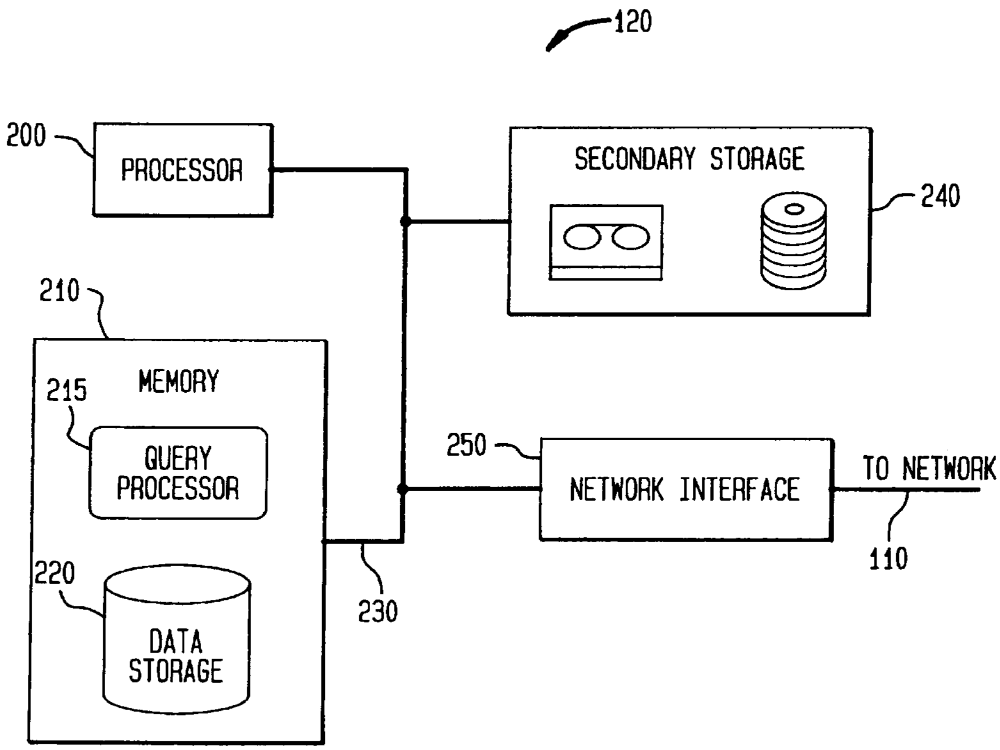

120
200
PROCESSOR
SECONDARY STORAGE
240
210
MEMORY
215
QUERY PROCESSOR
250
NETWORK INTERFACE
TO NETWORK
110
220
DATA STORAGE
230

FIG. 3

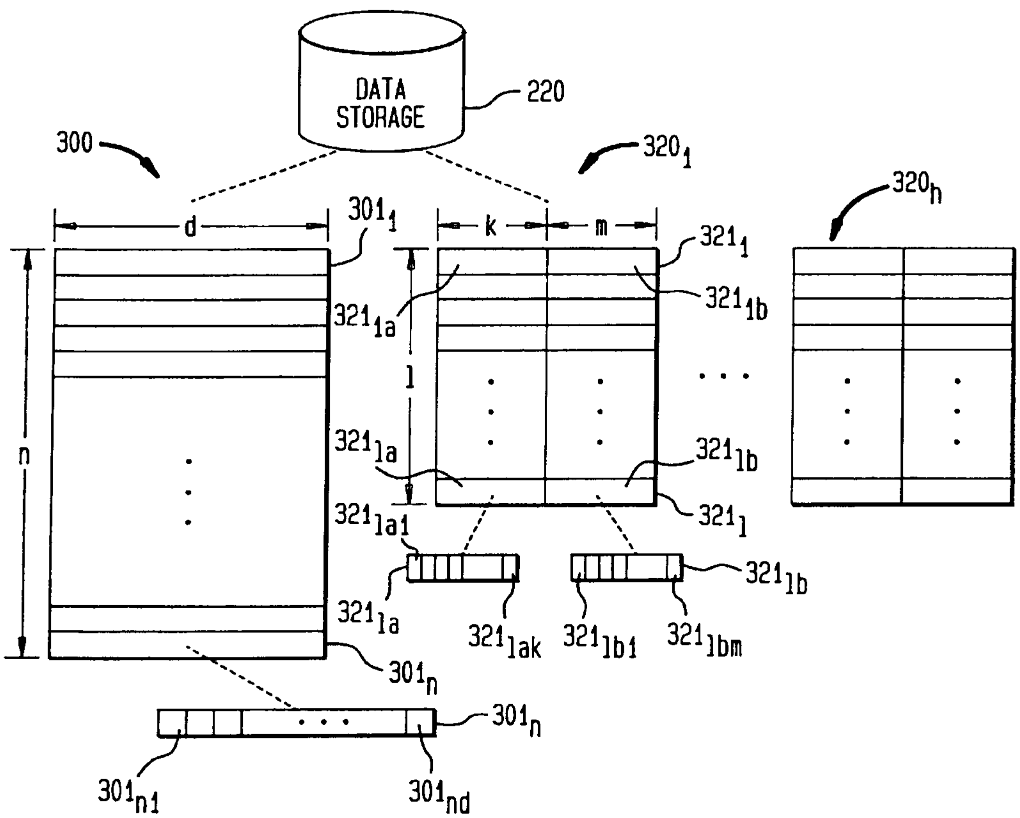

DATA
STORAGE
220
300
320_1
320_h
d
301_1
k
m
321_1
n
321_{1a}
321_{1b}
l
321_{la}
321_{lb}
321_l
321_{la1}
321_{lb}
321_{la}
321_{lak}
321_{lb1}
321_{lbm}
301_n
301_n
301_{n1}
301_{nd}

FIG. 4

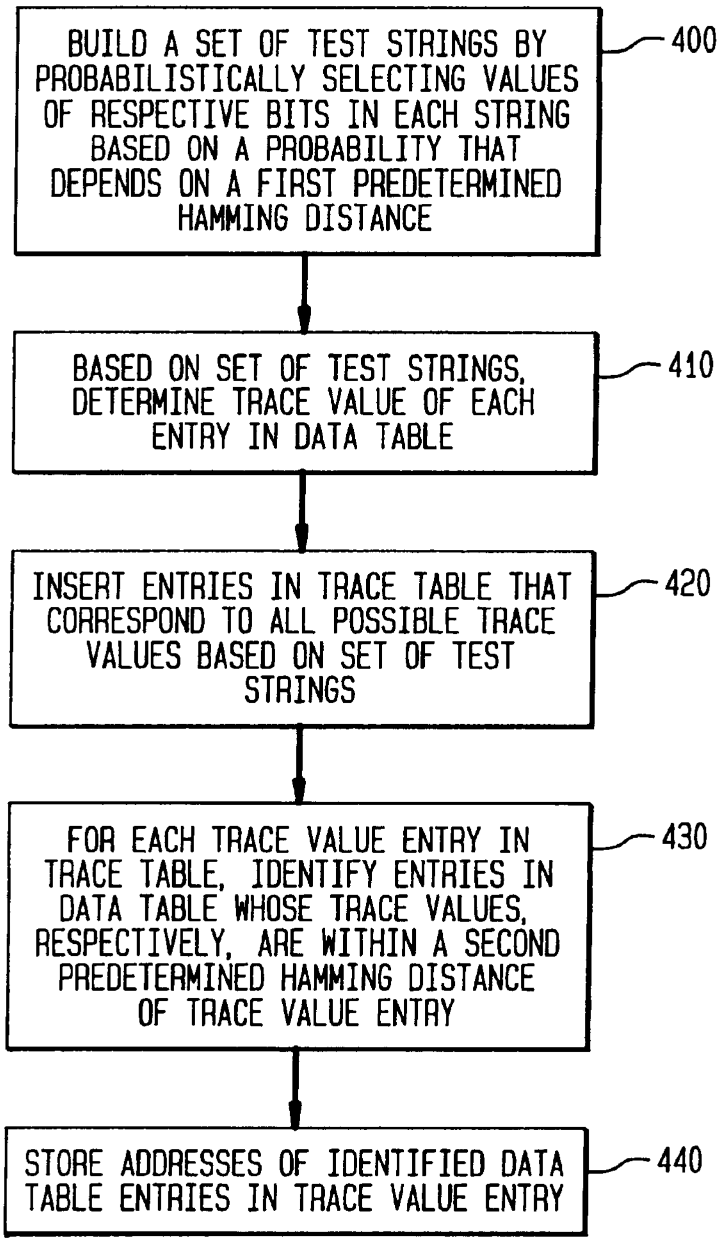

BUILD A SET OF TEST STRINGS BY PROBABILISTICALLY SELECTING VALUES OF RESPECTIVE BITS IN EACH STRING BASED ON A PROBABILITY THAT DEPENDS ON A FIRST PREDETERMINED HAMMING DISTANCE
400
BASED ON SET OF TEST STRINGS, DETERMINE TRACE VALUE OF EACH ENTRY IN DATA TABLE
410
INSERT ENTRIES IN TRACE TABLE THAT CORRESPOND TO ALL POSSIBLE TRACE VALUES BASED ON SET OF TEST STRINGS
420
FOR EACH TRACE VALUE ENTRY IN TRACE TABLE, IDENTIFY ENTRIES IN DATA TABLE WHOSE TRACE VALUES, RESPECTIVELY, ARE WITHIN A SECOND PREDETERMINED HAMMING DISTANCE OF TRACE VALUE ENTRY
430
STORE ADDRESSES OF IDENTIFIED DATA TABLE ENTRIES IN TRACE VALUE ENTRY
440

FIG. 5

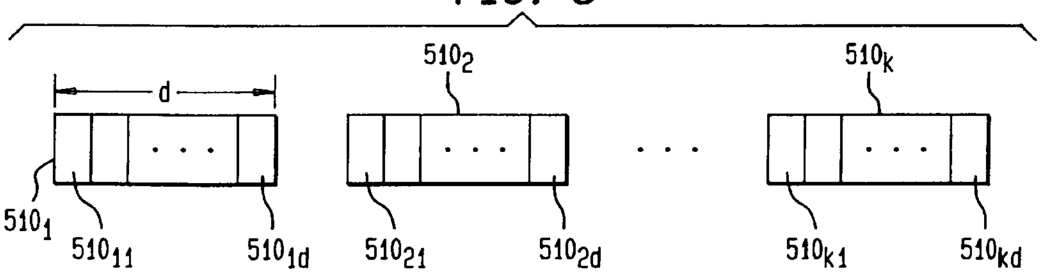


FIG. 6

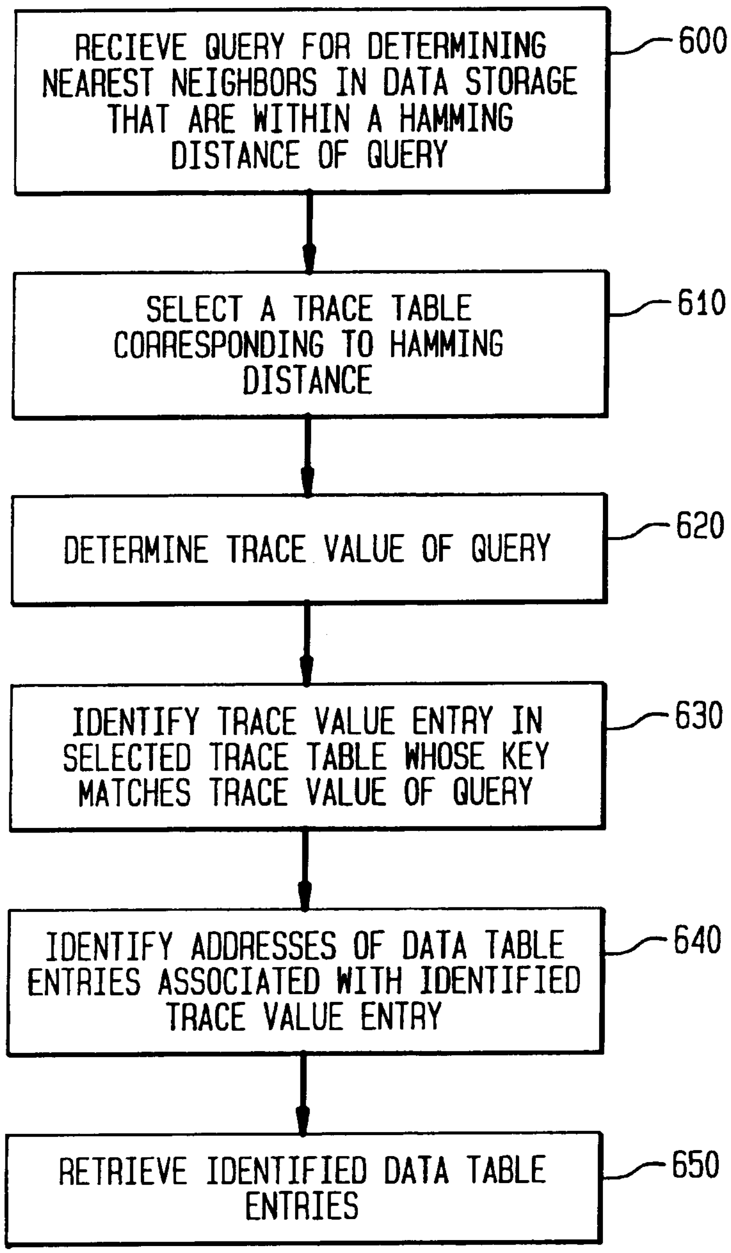
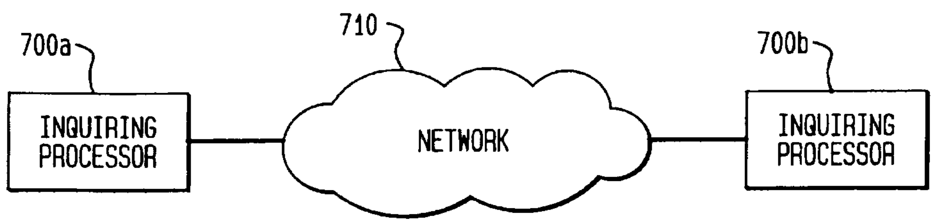

RECIEVE QUERY FOR DETERMINING NEAREST NEIGHBORS IN DATA STORAGE THAT ARE WITHIN A HAMMING DISTANCE OF QUERY
600
SELECT A TRACE TABLE CORRESPONDING TO HAMMING DISTANCE
610
DETERMINE TRACE VALUE OF QUERY
620
IDENTIFY TRACE VALUE ENTRY IN SELECTED TRACE TABLE WHOSE KEY MATCHES TRACE VALUE OF QUERY
630
IDENTIFY ADDRESSES OF DATA TABLE ENTRIES ASSOCIATED WITH IDENTIFIED TRACE VALUE ENTRY
640
RETRIEVE IDENTIFIED DATA TABLE ENTRIES
650

FIG. 7


700a
INQUIRING PROCESSOR
710
NETWORK
700b
INQUIRING PROCESSOR

METHOD FOR DETERMINING APPROXIMATE HAMMING DISTANCE AND APPROXIMATE NEAREST NEIGHBORS OF A QUERY

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Ser. No. 60/066,936, filed Nov. 17, 1997, the contents of which are hereby incorporated by reference.

BACKGROUND OF THE INVENTION

The present invention relates generally to information retrieval from electronic storage devices, and more particularly, to a method and system for determining approximate hamming distance of two strings and approximate nearest neighbors of a query.

Comparing files or documents that reside remotely in different inquiring processors in a network is a task, which requires significant communication between the inquiring processors. For example, when a first inquiring processor wishes to compare a first file that resides in the first inquiring processor with a second file that resides in a remote second inquiring processor, the first and second inquiring processors must communicate the files or information about the files over the network.

The least sophisticated method for determining whether the two files match each other is to transmit one of the files over the network and to compare the files at one of the inquiring processors. Communicating an entire file, of course, is not efficient since the size of the file may be large.

A more efficient method for comparing the two files is to communicate, for example, the hash value of one of the files over the network and to compare the respective hash values of the files at one of the inquiring processors. This method, however, only checks for an exact match between the two files.

Hence, it is desirable to estimate at an inquiring processor how closely two files match each other. A hamming distance is one measure of how closely two files or strings match each other. For example, given two strings that are of equal length and include a sequence of bits, the hamming distance of the two strings represents the number of non-matching bits in the two strings.

Similarly, in electronic storage applications, an entry in an electronic storage device is a nearest neighbor of a query when the content of that entry is the closest match from among other entries in the storage device. For example, if the query and the entries in the storage device each include a sequence of d bits, a nearest neighbor entry in the storage device is an entry that has the least number of non-matching bits when compared with the query.

Searching for entries that are the nearest neighbors of a query is a task, which is performed in a variety of applications, including information retrieval, data mining, web search engines and other web related applications, pattern recognition, machine learning, computer vision, data compression, and statistical analysis. Many of these applications represent the entries in an electronic storage device as vectors in a high dimensional space. For example, one known method for textual information retrieval uses a latent semantic indexing, where the semantic contents of the entries and queries are represented as vectors in a high dimensional space.

The least sophisticated method for searching an electronic storage device for the nearest neighbors of a query is to

compare, on-line or off-line, each entry in the storage device with the query. Comparing each and every entry with the query, of course, is not practical since the size of an average electronic device is large and continues to increase with the advancements in information storage technology.

Other known methods attempt to reduce the high dimensional representation of entries in electronic storage devices. For example, J. Kleinberg, "Two Algorithms For Nearest-Neighbor Search In High Dimensions," in the proceedings of 29th Symposium Of Theory Of Computing, pp. 599–608 (1997), discloses two algorithms for reducing the search space when determining the nearest neighbors in an electronic storage device. The Kleinberg algorithms search for the nearest neighbors by drawing random projections from vectors, which represent the entries in the storage device, to a set of random lines in Euclidean space.

P. Indyk and R. Motwani, "Approximate Nearest Neighbors: Towards Removing The Curse Of Dimensionality," in the proceedings of 30th Symposium Of Theory Of Computing (1998), discloses another algorithm for reducing the search space. The Indyk and Motwani algorithm searches for the nearest neighbors in an electronic storage device by partitioning the search space into spheres and by categorizing the entries in the storage device into buckets.

The above methods, however, requirement significant processing and storage resources. Therefore, it is desirable to have a method and system for overcoming the above and other disadvantages of the prior art.

DESCRIPTION OF THE INVENTION

Methods and systems consistent with the present invention determine whether a first string in an electronic storage device resides within a first hamming distance of a second string in the storage device. As used herein, "electronic storage device" refers to any processing system that stores information that a user at an inquiring processor may wish to retrieve. Moreover, the terms "electronic storage device" and "database" will be used interchangeably and should be understood in their broadest sense.

In one embodiment, a database determines one or more nearest neighbors of a query that are within a first hamming distance of the query. The database prebuilds a first set of strings by probabilistically selecting values of respective bits in each of the first set of strings based on a probability that depends on the first hamming distance. Based on the first set of strings, the database predetermines the trace values of data entries in the database, respectively, and stores the predetermined trace values as entries in a trace table.

For each trace value entry, the database identifies the data entries whose trace values are within a second hamming distance of the trace value entry, and stores the addresses of the identified data entries in the trace value entry. When the database receives a query, by identifying the trace value entry in the trace table that matches the trace value of the query, the database determines whether any data entries are within the first hamming distance of the query.

In another embodiment, a first processor communicates with a second processor to determine whether a first string that resides in the first processor is within a first hamming distance of a second string that resides in the second processor. The first processor and the second processor each have access to a shared third string that includes a plurality of bits, where the value of each bit is probabilistically pre-selected based on a probability that depends on a first hamming distance. The first processor computes a first inner product of the first string with the third string, and sends the

first inner product to the second processor. When the second processor receives the first inner product, the second processor computes a second inner product of the second string with the third string.

The second processor compares the first inner product with the second inner product to determine whether the first string is within the first hamming distance of the second string as follows: The second processor determines that the distance between the first string and the second string is less than the first hamming distance when the first inner product equals the second inner product. The second processor determines that the distance between the first string and the second string is greater than the first hamming distance when the first inner product is different from the second inner product.

In yet another embodiment, the first processor and the second processor each have access to a shared set of strings that include a plurality of bits, where the value of each bit is probabilistically pre-selected based on a probability that depends on a first hamming distance. The first processor computes a first set of inner products of the first string with each of the set of strings, and sends the first set of inner products to the second processor. When the second processor receives the first set of inner products, the second processor computes a second set of inner products of the second string with the each of the set of strings.

The second processor compares the first set of inner products with the second inner products to determine whether the first string is within the first hamming distance of the second string as follows: The second processor determines that the distance between the first string and the second string is less than a first hamming distance when the distance between the first set of inner products and the second set of inner products is less than a second predetermined hamming distance. The second processor determines that the distance between the first string and the second string is greater than the first hamming distance when the distance between the first set of inner products and the second set of inner products is greater than the second predetermined hamming distance.

The description of the invention and the following description for carrying out the best mode of the invention should not restrict the scope of the claimed invention. Both provide examples and explanations to enable others to practice the invention. The accompanying drawings, which form part of the description for carrying out the best mode of the invention, show several embodiments of the invention, and together with the description, explain the principles of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

In the Figures:

FIG. **1** is a block diagram of an inquiring processor connected to a database, in accordance with an embodiment of the invention;

FIG. **2** is a block diagram of a database, in accordance with an embodiment of the invention;

FIG. **3** is a block diagram of a data storage in a database, in accordance with an embodiment of the invention;

FIG. **4** is a flow chart of the steps performed by a query processor for configuring a trace table, in accordance with an embodiment of the invention;

FIG. **5** is a block diagram of a set of test strings for configuring a trace table, in accordance with an embodiment of the invention;

FIG. **6** is a flow chart of the steps performed by a query processor for determining the approximate nearest neighbors of a query, in accordance with an embodiment of the invention; and

FIG. **7** is a block diagram of a first inquiring processor communicating via a network with a second inquiring processor, in accordance with an embodiment of the invention.

BEST MODE FOR CARRYING OUT THE INVENTION

Reference will now be made in detail to the preferred embodiments of the invention, examples of which are illustrated in the accompanying drawings. Wherever possible, the same reference numbers will be used throughout the drawings to refer to the same or like parts.

FIG. **1** is a block diagram of an inquiring processor **100** connected via a network **110** to a database **120**, in accordance with an embodiment of the present invention. Inquiring processor **100** may comprise any form of computer capable of generating and transmitting data, for example a query. Inquiring processor **100** can be programed with appropriate application software to implement the methods and systems described herein.

Network **110** comprises any conventional communications network either internal or external for affecting communication between inquiring processor **100** and database **120**. Network **110** may comprise, for example, an internal local area network or a large external network, such as the Internet.

Database **120** includes any conventional data storage or any set of records or data, which are stored, for example, as bits. FIG. **2** is a block diagram of database **120**, in accordance with an embodiment of the present invention. Database **120** comprises processor **200** connected via bus **230** to a memory **210**, a secondary storage **240**, and a network interface card **250**, which interfaces network **110**. Memory **210** comprises a data storage **220** and a query processor **215**, which includes instructions in the form of software that processor **200** executes.

Secondary storage **240** comprises a computer readable medium such as a disk drive and a tape drive. From the tape drive, software and data may be loaded onto the disk drive, which can then be copied into memory **210**. Similarly, software and data in memory **210** may be copied onto the disk drive, which can then be loaded onto the tape drive.

FIG. **3** is a block diagram of a data storage **220**, in accordance with an embodiment of the invention. As shown, data storage **220** includes a data table **300** and a set of h trace tables $\mathbf{320}_1$ through $\mathbf{320}_h$, where h is an integer greater than zero. Data table **220** includes n entries $\mathbf{301}_1$ through $\mathbf{301}_n$, each of which includes a sequence of d bits, where n and d are also integers greater than zero. For example, as shown in FIG. **3**, entry $\mathbf{301}_n$ in data table **300** includes bits $\mathbf{301}_{n1}$ through $\mathbf{301}_{nd}$.

Trace tables $\mathbf{320}_1$–$\mathbf{320}_h$ correspond to a set of predetermined hamming distances, respectively. Each trace table $\mathbf{320}_1$–$\mathbf{320}_h$ includes l entries $\mathbf{321}_1$ through $\mathbf{321}_l$, each of which includes a trace value field and a data index field, where l is an integer greater than zero. For example, as shown, entry $\mathbf{321}_l$ in trace table $\mathbf{320}_1$ includes a trace value field $\mathbf{321}_{la}$ and a data index field $\mathbf{321}_{lb}$. Trace value field $\mathbf{321}_{la}$ includes k bits $\mathbf{321}_{la1}$, through $\mathbf{321}_{lak}$, where k is an integer greater than zero. Data index field $\mathbf{321}_{lb}$ includes m sub-fields $\mathbf{321}_{lb1}$ through $\mathbf{321}_{lbm}$, each of which includes, for example, the address of an entry in data table **300**, where m is an integer greater than zero.

FIG. 4 is a flow chart of the steps performed by query processor **215** for configuring, for example, trace table **320**$_1$, in accordance with an embodiment of the invention. Query processor **215** builds a set of k test strings **510**$_1$ through **510**$_k$ (step **400**), which are illustrated in FIG. **5**. Each test string **510**$_1$–**510**$_k$ includes a sequence of d bits. For example, as shown, test string **510**$_1$ includes bits **510**$_{11}$–**510**$_{1d}$. Query processor **215** probabilistically sets values of the bits in each test string **510**$_1$–**510**$_k$ independently at random based on a probability that depends on a first predetermined hamming distance H. Query processor **215** may predetermine the probability of setting a bit to 1 to be, for example, 1/(2H), and the probability of setting a bit to 0 to be 1−1/(2H).

Alternatively, in another embodiment, each entry **301**$_1$–**301**$_n$ in data table **300** and test string **510**$_1$–**510**$_k$ may include a sequence of d numbers, which are selected from a finite set of numbers that includes 0. Query processor **215** probabilistically selects the numbers in each test string **510**$_1$–**510**$_k$ based on a probability that depends on the first predetermined hamming distance H. In this embodiment, query processor **215** may predetermine the probability of selecting the number 0 to be, for example, 1−1/(2H), and the probability of selecting other numbers to be 1/(2H(d−1)).

Based on test strings **510**$_1$–**510**$_k$, query processor **215** determines trace values of entries **301**$_1$–**301**$_n$, respectively, in data table **300** (step **410**). Query processor **215** determines an inner product of each entry **301**$_1$–**301**$_n$ with each of test strings **510**$_1$–**510**$_k$. For example, for each entry **301**$_1$–**301**$_n$, query processor **215** identifies in the entry the bits that correspond to the 1 bits in test string **510**$_1$. Query processor **215** then performs an exclusive OR operation on the identified bits, the result of which is the first bit of the trace value associated with the entry. Query processor **215** then repeats this step using the remaining test strings **510**$_2$–**510**$_k$. Finally, query processor **215** builds a trace value associated with the entry by arranging in a sequence the resulting k bits from the k exclusive OR operations.

Alternatively, in an embodiment where each entry **301**$_1$–**301**$_n$ in data table **300** and test string **510**$_1$–**510**$_k$ include a sequence of d numbers, query processor **215** determines a vector product of each entry **301**$_1$–**301**$_n$. with each test string **510**$_1$–**510**$_k$. For example, for each entry **301**$_1$–**301**$_n$, query processor **215** multiplies each of the corresponding numbers in the entry with test string **510**$_1$ and sums the resulting d numbers modulo p, where p is an integer greater than zero. Query processor **215** then repeats this step using the remaining test strings **510**$_2$–**510**$_k$. Finally, query processor **215** builds a trace value associated with the entry by arranging in a sequence the resulting numbers based on test strings **510**$_1$–**510**$_k$.

Query processor **215** inserts into trace table **320**$_1$ l entries, which correspond to trace values that are based on test strings **510**$_1$–**510**$_k$ (step **420**). The number of entries l may be 2^k entries or all possible trace values corresponding to test strings **510**$_1$–**510**$_k$. Alternatively, the number of entries l may be a subset of all possible trace values.

For each trace value entry **321**$_1$–**321**$_l$ in trace table **320**$_1$, query processor **215** identifies the entries in data table **300** whose trace values (as determined in step **410**) are within a second predetermined hamming distance of the trace value entry (step **430**). For example, for each entry **301**$_1$–**301**$_n$ in data table **300**, query processor **215** determines whether the trace value associated with entry **301**$_1$–**301**$_n$ is within a second predetermined hamming distance of trace value entry **321**$_{1a}$ in trace table **320**$_1$. If the hamming distance between the trace value associated with entry **301**$_1$–**301**$_n$ and trace value entry **321**$_{1a}$ is less than or equal to the second predetermined hamming distance, query processor **215** stores the address of entry **301**$_1$–**301**$_n$ in data index field **321**$_{1b}$ (step **440**).

Finally, query processor **215** repeats steps **400**–**440** as described above for the remaining trace tables **510**$_2$–**510**$_h$ based on different sets of test strings that correspond to different predetermined hamming distances, respectively.

FIG. **6** is a flow chart of the steps performed by query processor **215** for determining the approximate nearest neighbors of a query transmitted by inquiring processor **100** to database **120**, in accordance with an embodiment of the invention. In this embodiment, query processor **215** receives from inquiring processor **100** a query, which includes a sequence of d bits (step **600**). Query processor **215** selects a trace table, for example trace table **320**$_1$, which is configured for a particular hamming distance (step **610**).

Query processor **215** then determines the trace value of the query based on the set of test strings **510**$_1$–**510**$_k$, which are associated with trace table **320**$_1$, as follows (step **620**): Query processor **215** determines an inner product of the query with each of test strings **510**$_1$–**510**$_k$. For example, query processor **215** identifies the bits in the query that correspond to the 1 bits in, for example, test string **510**$_1$. Query processor **215** then performs an exclusive OR operation on the identified bits. Query processor **215** then repeats this step using the remaining test strings **510**$_2$–**510**$_k$. Finally, query processor **215** builds a trace value associated with the query by arranging in a sequence the resulting bits from each exclusive OR operation.

From trace table **320**$_1$, query processor **215** identifies a trace value entry whose trace value field matches the trace value of the query (step **630**). Query processor **215** determines whether the data index field in the identified trace value entry includes addresses of one or more entries in data table **300** (step **640**).

If the data index field includes such an address, query processor **215** retrieves from data table **300** the identified entries, and sends the entries to inquiring processor **100**. Otherwise, using a binary search, query processor **215** selects from among trace tables **320**$_2$–**320**$_h$ a trace table that corresponds to a different hamming distance. Then, query processor **215** repeats steps **600**–**640** using the new trace table and associated test strings until query processor **215** identifies one or more entries in data table **300**.

FIG. 7 is a block diagram of an inquiring processor **700***a* connected via a network **710** to an inquiring processor **700***b*, in accordance with another embodiment of the invention. Inquiring processors **700***a* and **700***b* may each comprise any form of computer capable of generating and transmitting data, for example a query. Inquiring processors **700***a* and **700***b* can be programed with appropriate application software to implement the methods and systems described herein.

Network **710** comprises any conventional communications network either internal or external for affecting communication between inquiring processors **700***a* and **700***b*. Network **710** may comprise, for example, an internal local area network or a large external network, such as the Internet.

In one embodiment, inquiring processor **700***a* communicates with inquiring processor **700***b* to determine whether a first string that resides in inquiring processor **700***a* is within a first hamming distance H of a second string that resides in inquiring processor **700***b*. The first string and the second string each include a sequence of d bits. Furthermore,

inquiring processors **700***a* and **700***b* each have access to a shared test string that includes a sequence of d bits, where the value of each bit is probabilistically pre-selected at random based on a probability that depends on the first hamming distance H. The probability of selecting a bit to be a 1 bit may, for example, be 1/(2H), and the probability of selecting a bit to be a 0 bit may be 1−1/(2H).

Inquiring processor **700***a* computes a first inner product of the first string with the shared test string, and sends via network **710** the first inner product to inquiring processor **700***b*. When inquiring processor **700***b* receives the first inner product, inquiring processor **700***b* computes a second inner product of the second string with the shared test string.

Inquiring processor **700***b* compares the first inner product with the second inner product to determine whether the first string is within the first hamming distance H of the second string as follows: Inquiring processor **700***b* determines that the distance between the first string and the second string is less than the first hamming distance H when the first inner product equals the second inner product. Inquiring processor **700***b* determines that the distance between the first string and the second string is greater than the first hamming distance H when the first inner product is different from the second inner product.

Finally, inquiring processor **700***b* sends via network **710** the result of the comparison to inquiring processor **700***a*.

In another embodiment, the first string and the second string each include a sequence of d numbers. Furthermore, inquiring processors **700***a* and **700***b* each have access to a shared test string that includes a sequence of d numbers, where each number is probabilistically pre-selected from a set of finite numbers that includes the number 0 based on a probability that depends on a first hamming distance H. The probability of selecting the number 0 may, for example, be 1−1/(2H), and the probability of selecting the other numbers may be 1/(2H(d−1)).

Inquiring processor **700***a* computes a first vector product of the first string with the shared test string, and sends via network **710** the first vector product to inquiring processor **700***b*. When inquiring processor **700***b* receives the first vector product, inquiring processor **700***b* computes a second vector product of the second string with the shared test string.

Inquiring processor **700***b* compares the first vector product with the second vector product to determine whether the first string is within the first hamming distance H of the second string as follows: Inquiring processor **700***b* determines that the distance between the first string and the second string is less than the first hamming distance H when the first vector product equals the second vector product. Inquiring processor **700***b* determines that the distance between the first string and the second string is greater than the first hamming distance H when the first vector product is different from the second vector product.

Finally, inquiring processor **700***b* sends via network **710** the result of the comparison to inquiring processor **700***a*.

In yet another embodiment, to enhance the accuracy when determining whether a first string that resides in inquiring processor **700***a* is within a first hamming distance H of a second string that resides in inquiring processor **700***b*, inquiring processors **700***a* and **700***b* each have access to a shared set of k test strings. Each of the k test strings includes a sequence of d bits, where k<<d and the value of each bit is probabilistically pre-selected at random based on a probability that depends on the first hamming distance H. The probability of selecting a bit to be a 1 bit may, for example, be 1/(2H), and the probability of selecting a bit to be a 0 bit may be 1−1/(2H).

Inquiring processor **700***a* computes a first inner product of the first string with each of the k test strings, and sends via network **710** the first set of inner products to inquiring processor **700***b*. When inquiring processor **700***b* receives the first set of inner products, inquiring processor **700***b* computes a second inner product of the second string with each of the k test strings.

Inquiring processor **700***b* compares the first set of inner products with the second set of inner products to determine whether the first string is within the first hamming distance H of the second string as follows: Inquiring processor **700***b* determines that the distance between the first string and the second string is less than the first hamming distance H when the distance between first set of inner products and the second set of inner products is less than a second predetermined hamming distance. Inquiring processor **700***b* determines that the distance between the first string and the second string is greater than the first hamming distance H when the distance between the first set of inner products and the second set of inner products is greater than the second predetermined hamming distance.

Finally, inquiring processor **700***b* sends via network **710** the result of the comparison to inquiring processor **700***a*.

While it has been illustrated and described what are at present considered to be preferred embodiments and methods of the present invention, it will be understood by those skilled in the art that various changes and modifications may be made, and equivalents may be substituted for elements thereof without departing from the true scope of the invention.

In addition, many modifications may be made to adapt a particular element, technique or implementation to the teachings of the present invention without departing from the central scope of the invention. Therefore, it is intended that this invention not be limited to the particular embodiments and methods disclosed herein, but that the invention include all embodiments falling within the scope of the appended claims.

What is claimed is:

1. A method for determining whether a first string is within a first hamming distance of a second string, said method comprising the steps of:

building a third string that includes a plurality of bits, wherein the value of each bit depends on the first hamming distance;

computing a first inner product of the first string with the third string;

computing a second inner product of the second string with the third string; and

comparing the first inner product with the second inner product to determine whether the first string resides within the first hamming distance of the second string.

2. The method of claim **1**, wherein the building step comprises the step of:

probabilistically selecting the value of each bit in the third string based on a probability that depends on the first hamming distance.

3. The method of claim **1**, wherein the comparing step comprises the steps of:

determining that the distance between the first string and the second string is less than the first hamming distance when the first inner product equals the second inner product; and

determining that the distance between the first string and the second string is greater than the first hamming distance when the first inner product is different from the second inner product.

4. A method for identifying one or more entries in a database that are nearest neighbors of a query, wherein the identified entries are within a first hamming distance of the query, said method comprising the steps of:

building a first set of strings by selecting values of respective bits in each of the first set of strings based on the first hamming distance;

determining, based on the first set of strings, trace values of entries in the database, respectively;

determining, based on the first set of strings, a trace value of the query;

identifying the entries in the database whose trace values are within a second hamming distance of the determined trace value of the query.

5. The method of claim 4 further comprising the steps of:

determining a first set of trace values corresponding to the first set of strings; and

identifying, for each one of the first set of trace values, the entries in the database whose values are within a second hamming distance of the first set of trace values, respectively.

6. The method of claim 4, wherein the step of determining the trace values of the entries comprises the step of:

computing inner products of each of the first set of strings with each of the entries.

7. The method of claim 4, wherein the step of determining the trace value of the query comprises the step of:

computing inner products of each of the first set of strings with the query.

8. The method of claim 4, wherein the step of determining the trace values of the entries comprises the step of:

computing vector products of each of the first set of strings with each of the entries.

9. The method of claim 4, wherein the step of determining the trace value of the query comprises the step of:

computing vector products of each of the first set of strings with the query.

10. A method for determining whether a first string is within a first hamming distance of a second string, said method comprising the steps of:

building a third string that includes a plurality of numbers, wherein the value of each number depends on the first hamming distance;

computing a first vector product of the first string with the third string;

computing a second vector product of the second string with the third string; and

comparing the first vector product with the second vector product to determine whether the first string resides within the first hamming distance of the second string.

11. The method of claim 10, wherein the comparing step comprises the steps of:

determining that the distance between the first string and the second string is less than the first hamming distance when the first vector product equals the second vector product; and

determining that the distance between the first string and the second string is greater than the first hamming distance when the first vector product is different from the second vector product.

12. A method for determining whether a first string is within a first hamming distance of a second string, said method comprising the steps of:

building a set of strings that include a plurality of bits, wherein the value of each bit depends on the first hamming distance;

computing a first set of inner products of the first string with the set of strings;

computing a second set of inner products of the second string with the set of strings; and

comparing the first set of inner products with the second set of inner products to determine whether the first string resides within the first hamming distance of the second string.

13. The method of claim 12, wherein the comparing step comprises the steps of:

determining that the distance between the first string and the second string is less than the first hamming distance when the distance between the first set of inner products and the second set of inner products is less than a second predetermined hamming distance; and

determining that the distance between the first string and the second string is greater than the first hamming distance when the distance between the first set of inner products and the second set of inner products is greater than the second predetermined hamming distance.

14. A method for determining whether a first string is within a first hamming distance of a second string, said method comprising the steps of:

building a set of strings that include a plurality of bits, wherein the value of each bit depends on the first hamming distance;

computing a first set of vector products of the first string with the set of strings;

computing a second set of vector products of the second string with the set of strings; and

comparing the first set of vector products with the second set of vector products to determine whether the first string resides within the first hamming distance of the second string.

15. The method of claim 14, wherein the comparing step comprises the steps of:

determining that the distance between the first string and the second string is less than the first hamming distance when the distance between the first set of vector products and the second set of vector products is less than a second predetermined hamming distance; and

determining that the distance between the first string and the second string is greater than the first hamming distance when the distance between the first set of vector products and the second set of vector products is greater than the second predetermined hamming distance.

16. A computer-readable medium capable of configuring a database to perform a method for determining whether a first string is within a first hamming distance of a second string, said method comprising the steps of:

building a third string that includes a plurality of bits, wherein the value of each bit depends on the first hamming distance;

computing a first inner product of the first string with the third string;

computing a second inner product of the second string with the third string; and

comparing the first inner product with the second inner product to determine whether the first string resides within the first hamming distance of the second string.

17. The computer-readable medium of claim 16, wherein the comparing step comprises the steps of:

determining that the distance between the first string and the second string is less than the first hamming distance when the first inner product equals the second inner product; and

determining that the distance between the first string and the second string is greater than the first hamming distance when the first inner product is different from the second inner product.

* * * * *

X. RELATED PROCEEDINGS APPENDIX

None